CLAIMS

1. A method comprising:

capturing a web page from a web site;

extracting data from the web page using a data harvesting script;

normalizing the extracted data; and

storing the normalized data in a database.

2. A method as recited in claim 1 wherein the web site is associated with a financial institution.

3. A method as recited in claim 1 wherein the captured web page contains information regarding a customer's account at the financial institution.

4. A method as recited in claim 1 wherein the web page is an HTML screen.

5. A method as recited in claim 1 further comprising:

capturing a second web page from a second web site;

extracting data from the second web page using the data harvesting script;

normalizing the data extracted from the second web page; and

storing the normalized data from the second web page in the database.

6.   A method as recited in claim 1 further comprising generating an error message if data cannot be extracted from the web page using the data harvesting script.

7.   A method as recited in claim 1 further comprising storing a copy of the captured web page if data cannot be extracted from the web page using the data harvesting script.

8.   One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 1.

9.   A method comprising:

retrieving financial data associated with a user's financial account from a data source;

identifying data of interest retrieved from the data source;

normalizing the identified data; and

storing the normalized data in a database.

10.   A method as recited in claim 9 further comprising:

retrieving financial data associated with a user from a second data source;

normalizing the data retrieved from the second data source; and

storing the normalized data in the database.

11. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 9.

12. A method comprising:

capturing a web page from a web site;

attempting to extract data from the web page using a data harvesting script;

removing personal information from the captured web page;

storing the captured web page without the personal information; and

if data cannot be extracted from the web page, analyzing the web page and the data harvesting script to determine why data could not be extracted from the web page.

13. A method as recited in claim 12 further comprising editing the data harvesting script based on an analysis of the captured web page.

14. A method as recited in claim 12 further comprising:

editing the data harvesting script based on an analysis of the captured web page;

capturing a new version of the web page from the web site; and

extracting data from the web page using the edited data harvesting script.

15. A method as recited in claim 14 further comprising:

normalizing the data extracted from the web page; and

storing the normalized data in a database.

16.   One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 12.

17.   A method comprising:

capturing a first web page from a first financial institution web site;

capturing a second web page from a second financial institution web site;

extracting data from the first web page using a first data harvesting script;

extracting data from the second web page using a second data harvesting script;

normalizing the data extracted from the first web page and the second web page; and

storing the normalized data in a database.

18.   A method as recited in claim 17 further comprising generating an error message if data cannot be extracted from the first web page or the second web page.

19.   One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 17.

20.   A method as recited in claim 17 wherein capturing a first web page includes capturing a first set of web pages.

21.   A method as recited in claim 17 wherein capturing a second web page includes capturing a second set of web pages.

22.   An apparatus comprising:

a data capture module configured to capture a web page from a web site associated with a financial institution;

a data extraction module coupled to the data capture module and configured to extract data from the captured web page using a data harvesting script, the data extraction module further configured to normalize the extracted data; and

a database control module coupled to the data extraction module and configured to store the normalized data in a database.

23.   An apparatus as recited in claim 22 wherein the data capture module is further configured to retrieve financial data associated with a user's account from a data source.

24.   An apparatus as recited in claim 22 wherein the data extraction module is further configured to generate an error message if data cannot be extracted from the web page using the data harvesting script.

25. One or more computer readable media having stored thereon a plurality of instructions that, when executed by a processor, causes the processor to perform acts comprising:

capturing a web page from a financial institution web site;

attempting to extract data from the captured web page using a data harvesting script;

removing personal information from the captured web page;

storing the captured web page without the personal information; and

if data cannot be extracted from the web page, analyzing the web page to determine why data could not be extracted from the web page.

26. One or more computer readable media as recited in claim 25, wherein if data cannot be extracted from the web page, editing the data harvesting script based on an analysis of the captured web page.

27. One or more computer readable media as recited in claim 25, further comprising:

normalizing the data extracted from the web page; and

storing the normalized data in a database, wherein the database contains data extracted from other web pages.